Predicting Soccer Player Overall Rating: A Machine Learning Approach

Ramez Yousri, Yehia Kareem, Youstina Sherif, Zeyad Essam

Under the supervision of Dr.Esraa Afify
Faculty of Engineering, Software Engineering Department, Egyptian Chinese University

Abstract

This paper explores the application of supervised and unsupervised machine learning methodologies in predicting the overall rating scores of football players. Initially, a regression model was developed to forecast the continuous values of overall ratings, yielding a mean squared error of 3.482 and an R2 score of 0.92. Subsequently, employing the K-means clustering algorithm facilitated the segmentation of rating scores into distinct clusters. These clusters were then subjected to classification via a Decision Tree Classifier, achieving a notable accuracy of 99.7%. Through this analysis, the study contributes to the field of sports analytics by offering insights into the effectiveness of machine learning techniques in assessing the performance of football players.

Keywords

Soccer prediction, Machine Learning, Decision Tree, Unsupervised learning

1. Introduction

1.1 Background and Motivation

In recent years, the field of sports analytics has experienced a paradigm shift, owing to advancements in machine learning (ML) techniques and the availability of extensive player performance data(Apostolou and Tjortjis 2019) (Yiğit, Samak, and Kaya 2020) (Pantzalis and Tjortjis 2020). One of the fundamental aspects in assessing the prowess of football players is their overall rating. Overall rating is a composite measure encompassing various attributes reflecting their skill, performance, and impact on the game. Traditionally, the determination of these ratings has relied heavily on expert judgment and subjective evaluation.(ABİDİN 2021)

However, with the proliferation of data-driven approaches, there arises an opportunity to augment and refine the process of player assessment using ML methodologies. By harnessing the power and objective analysis of supervised and unsupervised learning techniques, it becomes feasible to develop predictive models capable of accurately estimating overall rating scores based on objective performance metrics(ABİDİN 2021).

1.2 Problem Statement

This paper endeavors to explore the efficacy of such ML approaches in the context of predicting overall rating scores for football players. Initially, a regression model is constructed to forecast the continuous values of overall ratings, providing insights into the relationship between player attributes and their respective ratings. Subsequently, leveraging clustering algorithms enables the categorization of rating scores into distinct groups, facilitating a deeper understanding of player performance profiles. Moreover, employing classification techniques further refines the analysis by assigning players to specific rating clusters with a high degree of accuracy.

1.3 Objectives

- Develop a regression model to forecast the continuous values of overall rating scores for football players.
- Assess the effectiveness of the regression model in predicting the overall outcome, thereby providing insights into the relationship between player attributes and their respective ratings.
- Utilize clustering algorithms to categorize rating scores into distinct groups, facilitating a deeper understanding of player performance profiles.
- Create a classifier to classify the overall player rating into clusters, aiding coaches in approximating player scores and enhancing tactical decision-making processes.
- Achieve a high degree of accuracy in classifying players into specific rating clusters, demonstrating the effectiveness of machine learning techniques in objective player assessment.
- Contribute to the advancement of sports analytics by offering insights into predictive modeling, clustering, and classification of overall rating scores, thereby paving the way for more data-driven approaches to player evaluation and team management in football.

1.4 Contribution

This study makes several significant contributions to the field of sports analytics, particularly in the domain of assessing the performance of football players using machine learning methodologies.

Firstly, the research demonstrates the efficacy of supervised machine learning techniques, specifically regression modeling, in predicting the overall rating scores of football players. By developing a regression model, the study provides insights into the intricate relationship between various player attributes and their overall ratings. This not only enhances the objectivity, which in other methods can be affected by the subjectivity of the expert, of player assessment but also lays the groundwork for more data-driven and precise evaluation methods in football analytics.

Secondly, the application of unsupervised learning techniques, particularly K-means clustering, offers an unsupervised approach to segmenting rating scores into distinct clusters. This segmentation provides a deeper understanding of player performance profiles by identifying patterns and similarities among players based on their overall rating scores. Such insights can be helpful for coaches and talent scouts in identifying players with similar skill sets or playing styles, thus leading to more strategic team building and better decision-making.

Moreover, the study extends the analysis by employing classification algorithms, such as the decision tree classifier, to further refine the segmentation of rating scores into clusters. This classification not only aids coaches in approximating player scores but also enhances tactical decision-making by providing a more nuanced and abstract understanding of player capabilities and roles within the team.

2. Literature Review

Machine Learning (ML) found numerous applications within various fields from sports to medicine to entertainment. This led to a plethora of AI-related papers discussing its use in team sports e.g. Basketball and Football, and single sports like Tennis.

2.1. Artificial Intelligence

Defining artificial intelligence (AI) poses a significant challenge due to the lack of a universally accepted definition(H. Sheikh, Prins, and Schrijvers 2023). Various interpretations exist, ranging from equating AI with algorithms to describing it as the imitation of human intelligence by computers. While some definitions focus on the technology's ability to imitate complex human skills, others emphasize the performance of tasks in complex environments with a degree of autonomy(Kok et al. 2009). Despite attempts to articulate AI's essence through task-based definitions, ambiguity persists due to the broad scope of the concept and the elusive nature of human intelligence. Scholars acknowledge the difficulty in defining AI precisely, given its status as an imitation of human cognition, a domain still incompletely understood by researchers from diverse disciplines. Until a consensus is reached on the nature of human intelligence, clarity in defining and understanding AI will remain elusive.(Boden 2018).

There are multiple use cases or rather subfields of artificial intelligence these include(Yousri and Hany 2022):

1. **Machine Learning (ML):** Involves teaching an agent to do a particular task, through training on certain collected data

2. **Computer Vision:** involves categorizing images and videos to derive valuable insights, such as identifying faces.

3. **Natural Language Processing(NLP):** aids in comprehending and interpreting languages like French and English, extracting pertinent information from textual data, such as news articles.

4. **Deep Learning (DL):** represents an advanced form of learning that allows models to fit data points into intricate curves, enabling complex analyses.

5. **Reinforcement Learning (RL):** is a subset of AI where agents learn by experimenting and adjusting their actions based on trial and error.

2.2. Machine Learning (ML)

Arthur Samuel, renowned for his checkers-playing program, defined machine learning as the field enabling computers to learn without explicit programming. In essence, machine learning empowers machines to handle data efficiently, particularly when extracting information is challenging through conventional means. With the exponential growth of datasets, the demand for machine learning has surged across industries, for example, it's prevalent in marketing (Duarte, Zuniga-Jara, and Contreras 2022) (Ngai and Wu 2022), healthcare(Alanazi 2022) (Callahan and Shah 2017), driving efforts to enhance machines' autonomous learning capabilities. Mathematicians and programmers explore diverse approaches to enable machines to learn independently from vast datasets.(Pyda and Kareenhalli 2018) Notably, machine learning relies on a variety of algorithms tailored to specific data problems. Data scientists emphasize the absence of a universal algorithm, with the choice depending on the problem's nature, variables involved, and suitable models.

2.2.1. Supervised Learning

One prominent area within machine learning is supervised learning, where algorithms learn from labeled training data to map inputs to outputs. This process involves dividing the dataset into training and testing sets, with algorithms discerning patterns from the former to predict or classify outcomes in the latter. There are various algorithms for supervised learning, examples include decision trees(Charbuty and Abdulazeez 2021) , and support vector machines (SVM)(Suthaharan 2016).

2.2.2 Unsupervised Learning

Unsupervised learning emerges as a pivotal area, distinguished by algorithms that uncover hidden patterns and structures within unlabeled data. Unlike supervised learning, where datasets are labeled with corresponding outputs, unsupervised learning tasks involve exploring raw data without explicit guidance, aiming to unveil inherent relationships and clusters. A fundamental aspect of unsupervised learning involves the exploration and segmentation of data into meaningful groups or clusters based solely on intrinsic characteristics. This process does not rely on predefined labels or outcomes, allowing algorithms to autonomously identify similarities and differences among data points. In the literature, unsupervised learning garners significant attention for its versatility and applicability across various domains, including anomaly detection(Chandola, Banerjee, and Kumar 2009), customer segmentation(Cooil, Aksoy, and Keiningham 2008), and dimensionality reduction(Sorzano, Vargas, and Montano 2014).

2.3. Football Player Evaluation

In (McIntosh, Kovalchik, and Robertson 2019),The authors examined the role of subjective and objective evaluations in player performance assessment within the Australian Football League (AFL). They aimed to determine the extent to which performance indicators explain subjective ratings and compare subjective and objective assessments. Using Inside Football Player Ratings (IFPR) and AFL Player Ratings, alongside performance indicators, player role classification, age, and match outcomes, they applied linear mixed model and regression tree analyses. Results showed that kicks and handballs were key factors influencing subjective ratings, with models explaining IFPR accurately in a significant proportion of instances. The study highlights the importance of integrating both subjective and objective assessments for a comprehensive understanding of player performance in AFL organizations. In (Hadley et al. 2000),The authors contribute to the sports economic literature by shifting the focus from Major League Baseball to analyzing football production in the National Football League (NFL). Employing the Poisson regression model, they assess the performance of NFL teams and head coaches, basing their measure on a production process converting player skills into games won. Their findings underscore the significance of quality coaching in the production process, suggesting that efficient coaching can contribute to an additional three to four victories in a season

2.4. Football Player Evaluation using ML

In (Rommers et al. 2020),The study aimed to assess injury risk among elite youth football players using anthropometric and performance measures with a machine learning model. A cohort of 734 players from seven Belgian academies was followed for one season. Extreme gradient boosting algorithms predicted injury and classified injuries as overuse or acute based on preseason tests. Results showed half of the players sustained injuries, with the algorithm achieving 85% precision, recall, and accuracy in identifying injuries and 78% precision, recall, and accuracy in classifying them. The findings suggest machine learning can effectively predict and differentiate between injury types, aiding injury risk management in elite youth football. In (Manish., Bhagat, and Pramila 2021),In modern football, winning is paramount, necessitating accurate player evaluation to identify weaknesses and enhance performance. Manual evaluation is prone to errors and time-consuming. This study proposes a statistical model to predict player statistics based on previous session data, considering various game aspects. While literature suggests machine learning and deep learning for player performance prediction, none consider player positions. Thus, this study designs separate models for each position to predict performance accurately. Various machine learning and deep learning models were implemented and compared, with each position incorporating position-specific variables. The study aims to clarify model performance through comparative analysis.

3. Dataset and Preprocessing

3.1 Dataset Description

We used the Football Players Data dataset from Kaggle (Masood Ahmed, n.d.) as the main dataset for our models. This comprehensive dataset provides detailed information on approximately 17,000 FIFA football players, meticulously sourced from SoFIFA.com. It encompasses a wide array of player-specific data points. As mentioned by the authors, it can be used for various tasks first of which is player evaluation. Other tasks include market value assessment and wage prediction, team composition and strategy planning, and machine learning models to predict future player potential and career trajectories.

Column Name	Description
Full Name	Player's full name.
Birth Date	Player's date of birth.
Age	Player's age.
Height (cm)	Player's height in centimeters.
Weight (kg)	Player's weight in kilograms.
Positions	Positions the player can play.
Nationality	Player's nationality.
Overall Rating	Player's overall rating in FIFA.
Potential	Player's potential rating in FIFA.
Value (Euro)	Market value of the player in euros.
Wage (Euro)	Weekly wage of the player in euros.
Preferred Foot	Player's preferred foot.
International Reputation (1-5)	International reputation rating from 1 to 5.
Weak Foot (1-5)	Rating of the player's weaker foot from 1 to 5.
Skill Moves (1-5)	Skill moves rating from 1 to 5.
Body Type	Player's body type.
Release Clause (Euro)	Release clause of the player in euros.

National Team	National team of the player.
National Rating	Rating in the national team.
National Team Position	Position in the national team.
National Jersey Number	Jersey number in the national team.
Various Skill Ratings	Ratings for specific skills such as crossing, finishing, heading accuracy, etc.

Table 1: Player attributes

3.2 Analysis

This visualization presents the distribution of player ages using a histogram: the x-axis represents age intervals, while the y-axis shows the frequency or count of players within each age interval. The histogram bars are colored blue, and the title "Distribution of Player Ages" provides an overview of the plot's purpose [Fig 1].

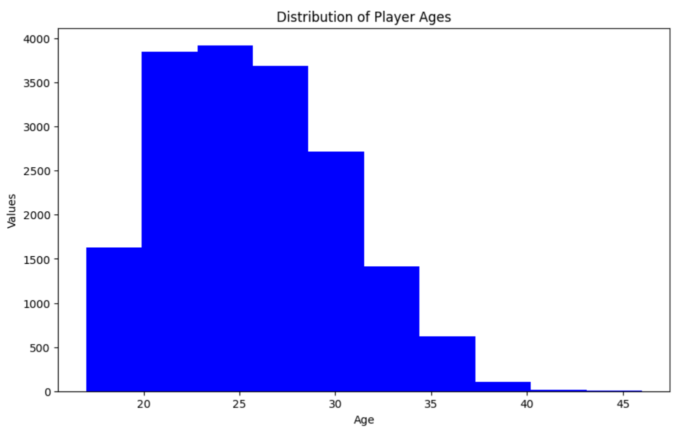

fig1: Distribution of Player Ages

This visualization presents the top 10 nationalities of FIFA players through a bar plot the x-axis represents the nationality labels, while the y-axis indicates the count of players for each nationality each bar's height corresponds to the number of players from a specific nationality the plot uses a color palette called 'viridis' for visual distinction, and the title "Top 10 Nationalities of FIFA Players" provides context for the data the x-axis labels are rotated by 45 degrees for better readability

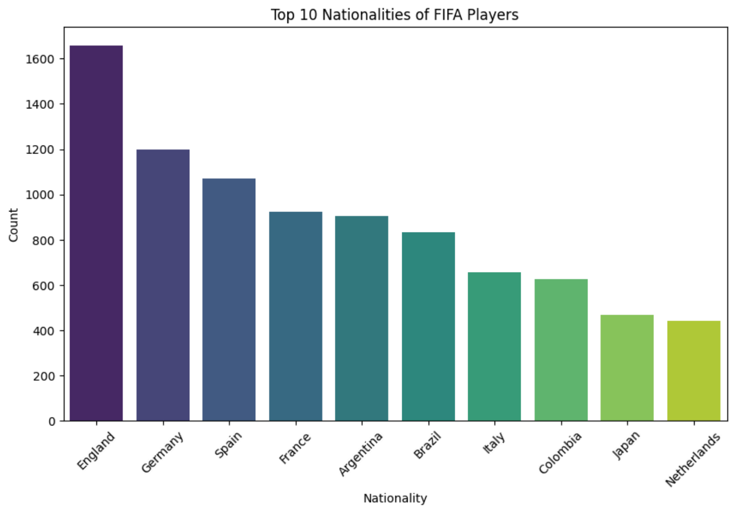

fig2: Top 10 nationalities

This visualization illustrates the distribution of overall ratings among FIFA players based on their preferred foot the x-axis represents the preferred foot (left or right), while the y-axis indicates the overall rating the width of the violin plots indicates the density of player ratings, with wider sections indicating higher density the plot's title, "Distribution of Overall Ratings by Preferred Foot," summarizes its purpose

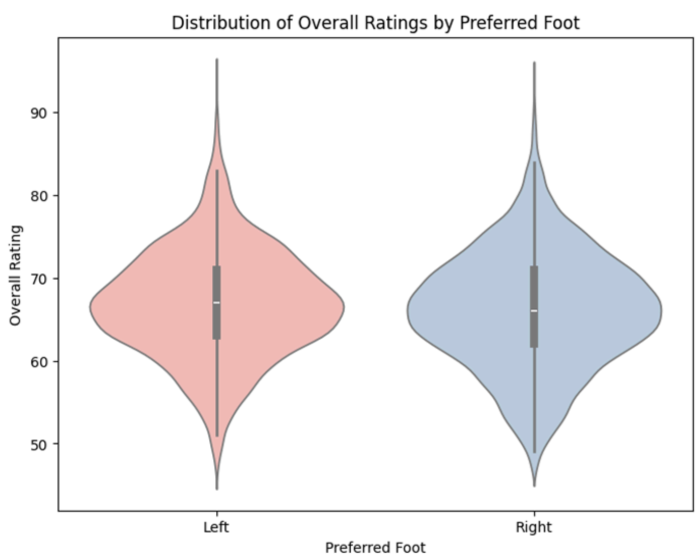

fig3: Distribution of overall ratings

This visualization depicts the distribution of international reputation among FIFA players the x-axis represents the international reputation levels (ranging from 1 to 5), while the y-axis indicates the count of players for each reputation level the bars' colors provide a visual distinction, with warmer colors indicating higher reputation levels. the title, "Distribution of International Reputation," summarizes the plot's content

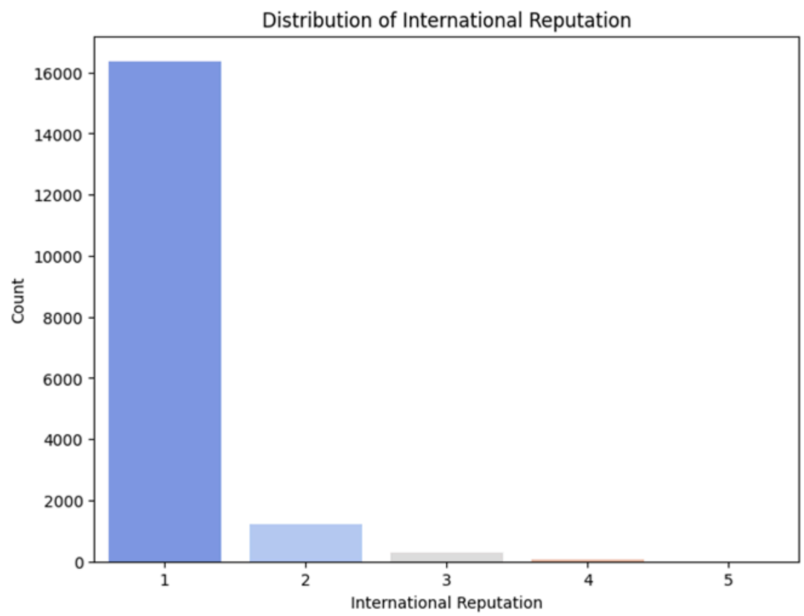

fig4:Distribution of international reputation

This visualization presents the distribution of preferred foot among FIFA players using a pie chart each slice of the pie represents a preferred foot category (left or right), with its size proportional to the percentage of players preferring that foot. The title, "Preferred Foot Distribution," summarizes the plot's content.

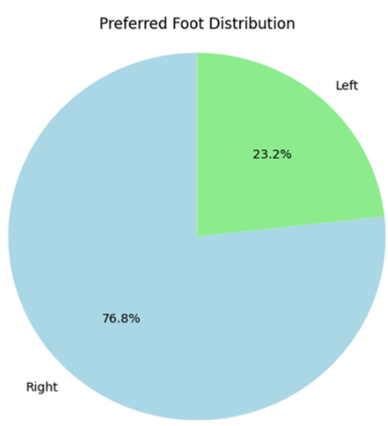

fig5: Preferred foot

This visualization compares the height and weight of FIFA players the blue line represents the height of players, while the green line represents their weight each point on the lines corresponds to a player's index, showing how their height and weight vary across the dataset the legend helps distinguish between the height and weight lines.

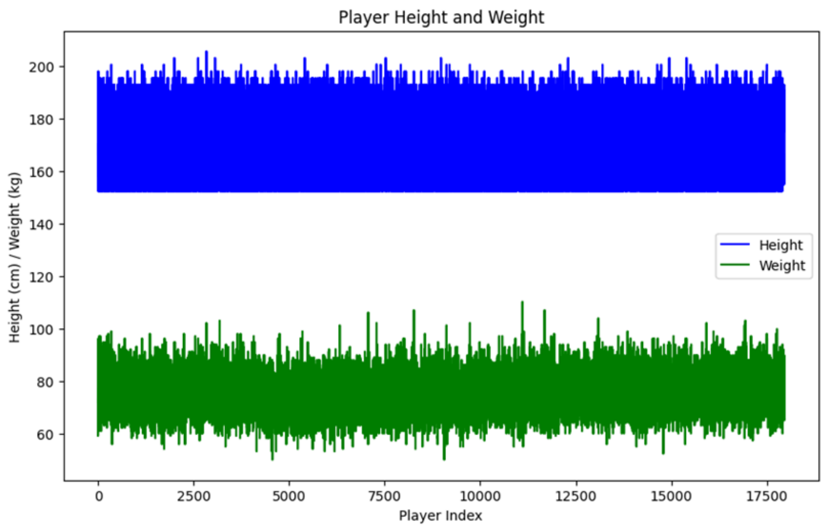

fig6: Player Height and weight

This visualization depicts the joint distribution of age and overall rating of FIFA players using a hexbin plot the x-axis represents the age of players, the y-axis represents their overall rating, and the density of points is represented by color intensity the title, "Joint Distribution of Age and Overall Rating," summarizes the plot's content.

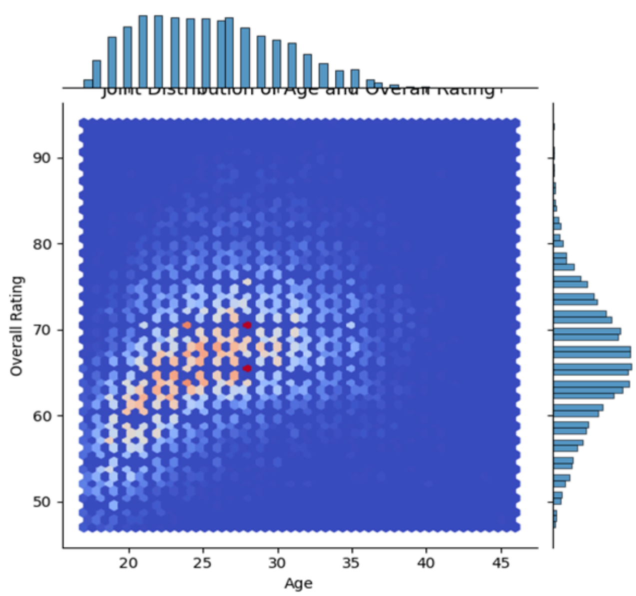

fig7: Joint diagram

This visualization displays the top 10 FIFA players ranked by their market value in euros each bar represents a player's market value, with the player's name shown on the x-axis and their corresponding market value displayed on the y-axis the bars are colored using a blue color palette, and the title "Top 10 Players by Market Value" summarizes the plot's content the x-axis labels are rotated by 45 degrees for better readability

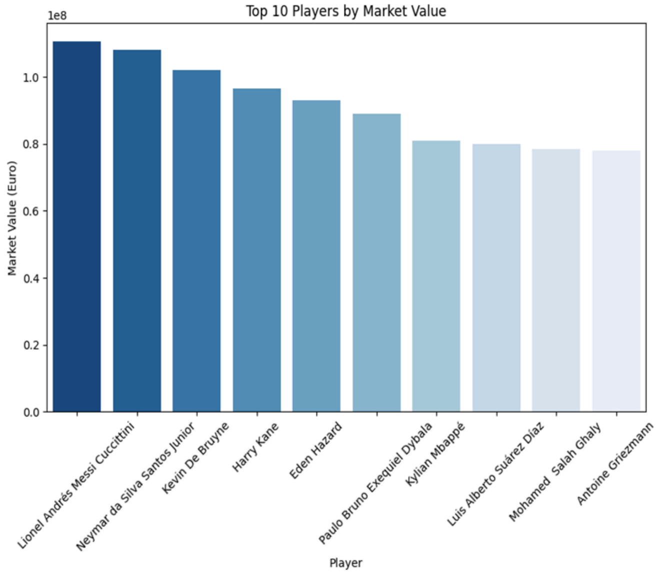

fig8: Top 10 Players by Market Value

This visualization illustrates the average potential of FIFA players across different ages the x-axis represents the players' ages, while the y-axis indicates their average potential each point on the line plot represents the mean potential of players at a specific age. The title, "Player Potential by Age," summarizes the plot's content.

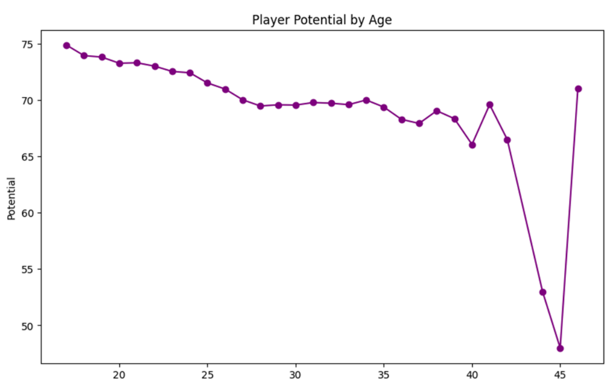

fig9: Player Potenial vs age

This visualization presents the distribution of player heights in centimeters using a histogram the x-axis represents height intervals, while the y-axis indicates the count of players falling within each interval the bars of the histogram are colored sky blue, and the title "Player Heights Distribution" provides context for the plot

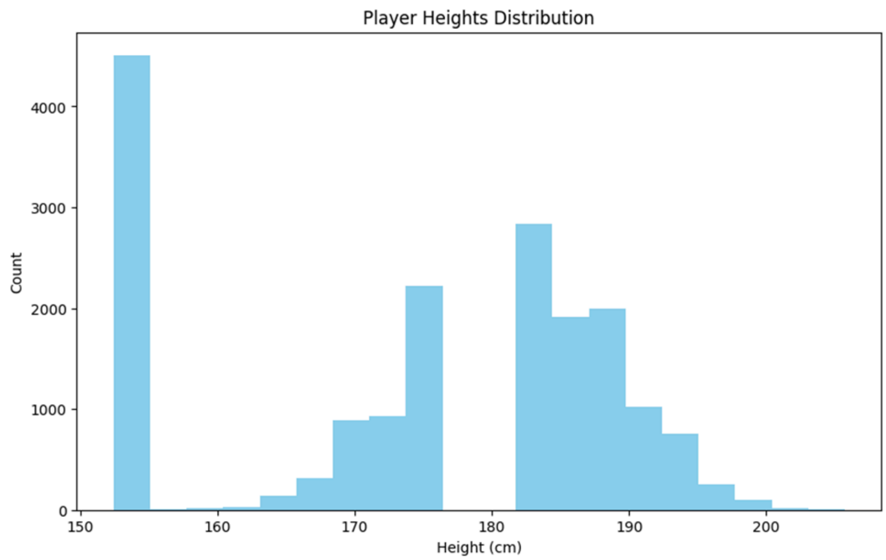

fig10: Player heights distribution

This visualization categorizes FIFA players into different market value ranges and then calculates the average potential for players within each range the x-axis represents these value ranges in euros, while the y-axis indicates the average potential of players within each range each point on the line plot corresponds to the mean potential of players within a specific value range the title, "Player Potential by Value Range," summarizes the plot's content.

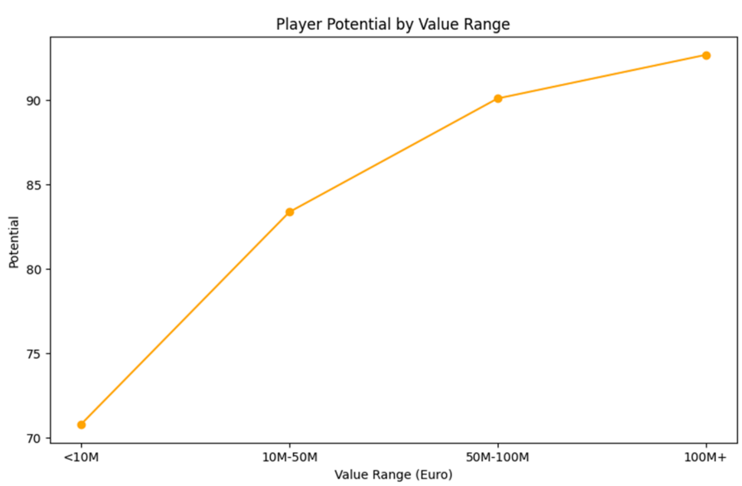

fig10: Player potential vs value

3.3 Data Preprocessing

In preparation for analysis and modeling, the dataset went through several preprocessing steps aimed at enhancing its quality and relevance. The initial phase involved handling missing values, ensuring that the dataset's integrity remained intact. Subsequently, a systematic feature selection process was undertaken to streamline the dataset for optimal predictive modeling. This involved removing columns that we believe are irrelevant to the prediction of overall rating scores. Specifically, attributes such as full name, birth date, body type, jersey number, national rating, national team position, and national jersey number were excluded from the dataset, as they were not considered influential factors in determining player performance.

Moreover, for categorical variables such as position and nationality, we used encoding methods like the one-hot encoding technique to facilitate their integration into the predictive models. However, recognizing the diversity of playing roles within football, a nuanced approach was adopted for encoding player positions. The football field was divided into four primary positions: attacker, center, defender, and goalkeeper, with each player's position mapped to one of these categories based on their predominant playing role. Similarly, nationalities were stratified into continents to provide a broader geographical context for analysis.

Furthermore, efforts were made to address redundant features within the dataset to improve computational efficiency without compromising on information integrity. For instance, both the age and birth date of players were included in the original dataset. However, given that age serves as a direct representation of a player's stage in their career, the decision was made to retain only the age column while discarding the redundant birth date column. This not only reduced computational overhead but also simplified the dataset structure, making it more conducive to analysis and modeling tasks.

4. Methodology

4.1 Models Used

The methodology employed in this study was designed to comprehensively assess the performance of football players while providing valuable insights for coaches and analysts. The approach consisted of a series of interconnected steps, each aimed at leveraging different machine-learning techniques to extract meaningful information from the dataset. To initiate the analysis, a regression model was used as the primary tool for predicting the continuous value of the overall score for each football player. This regression model was trained on a diverse array of player-specific attributes, including but not limited to age, position, skill ratings, and physical characteristics. By harnessing the power of regression analysis, the model provided a quantitative estimation of player performance, offering valuable insights into the factors contributing to overall rating scores.

In addition to the regression model, an unsupervised learning technique, specifically the K-means clustering algorithm, was employed to enhance the classification of players based on similarities in their overall rating scores. This clustering algorithm segmented the player population into distinct groups, with the number of clusters (num_clusters =3) predetermined to three. By grouping players into clusters, this approach provided a more abstract classification of player performance, allowing for the identification of common patterns and characteristics among players with similar overall ratings.

Furthermore, the clusters generated by the K-means algorithm served as the basis for training a Decision Tree classifier model. The classifier model utilized the clustered data to classify players into specific categories, thus offering a combined approach to measuring player performance. By integrating both regression-based predictions and cluster-based classifications, this hybrid model provided a comprehensive framework for evaluating player performance across different dimensions.

Moreover, the incorporation of a Decision Tree classifier enabled the extraction of actionable insights from the clustered data, offering football coaches and analysts valuable information for strategic decision-making.

5.Results

5.1.Regression Model

The regression model exhibited an initial mean squared error (MSE) score of 3.482, indicating a strong predictive performance. Upon normalization of the data, the MSE decreased substantially to 0.0016, and an R2 score of 0.930, further enhancing the model's accuracy and precision in predicting the overall rating scores of football players. This reduction in MSE highlights the efficacy of normalization techniques in refining the model's predictive capabilities, ultimately leading to improved performance metrics. Overall, these results underscore the effectiveness of the regression model in accurately forecasting overall rating scores based on player attributes and performance metrics.

5.2 K-Means Clustering

The K-means clustering algorithm, configured with several clusters equal to three, yielded promising performance metrics. The silhouette score, indicating the cohesion and separation of clusters, was notably high at 0.831. Additionally, the Calinski-Harabasz score, which measures the ratio of between-cluster dispersion to within-cluster dispersion, demonstrated strong clustering with a score of 32022.36. Furthermore, the Davies-Bouldin score, assessing the average similarity between each cluster and its most similar cluster, was calculated at 0.501, indicating well-separated clusters. Contrastingly, the random score and Fowlkes-Mallows score, serving as benchmarks for cluster evaluation, were comparatively lower at 0.213 and 0.215, respectively. These results collectively affirm the effectiveness of the K-means clustering algorithm in segmenting rating scores of football players into distinct clusters, facilitating a deeper understanding of player performance profiles.

5.3. Decision Tree Classifier

The Decision Tree classifier, trained on clusters as the target variable, demonstrated exceptional performance metrics. With an accuracy of 0.997, precision of 0.997, recall of 0.996, and F1 score of 0.996, the classifier exhibited robust predictive capabilities in accurately assigning football players to specific rating clusters. These high-performance metrics underscore the effectiveness of the Decision Tree classifier in accurately classifying players based on their performance profiles, thereby providing valuable insights for coaches and analysts in player assessment and strategic decision-making. you can find the confusion matrix in Fig.

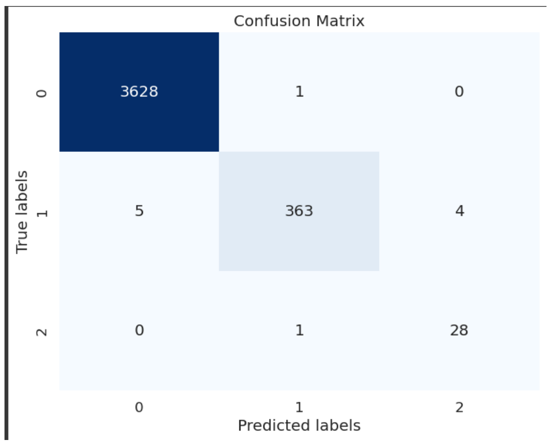

fig11. Confusion matrix of the decision tree classifier

Model	***Metric***	***Value***
Regression (Normalized)	*MSE*	0.0016
	R2	0.930
Decision Tree Classifier (differences occur after the 3rd decimal)	*Accuracy*	*0.972*
	Precision	*0.973*
	F1	*0.972*
	Recall	*0.972*
K-means	*Silhouette*	*0.83*
	CH	*32022.36*
	DB	*0.5010*
	Rand	*0.212*
	FM	*0.215*

Table2 : Summary of the results

6.Discussion

The findings presented in this study underscore the effectiveness of machine learning techniques in assessing and predicting the overall rating scores of football players. Beginning with the regression model, the substantial reduction in mean squared error (MSE) from 3.482 to 0.0016 after normalization highlights the importance of preprocessing techniques in enhancing predictive accuracy. The high R2 score of 0.930 further confirms the model's ability to capture variance in player attributes and performance metrics, emphasizing its utility in forecasting overall rating scores with precision.

Moving on to the K-means clustering algorithm, the robust performance metrics obtained, including a high silhouette score of 0.831 and a strong Calinski-Harabasz score of 32022.36, validate the efficacy of clustering techniques in segmenting football player rating scores into cohesive groups. The well-separated clusters, as indicated by the low Davies-Bouldin score of 0.501, emphasize the algorithm's ability to discern distinct performance profiles among players. However, the comparatively lower random and Fowlkes-Mallows scores suggest some limitations in cluster evaluation, warranting further exploration into cluster validation techniques.

Furthermore, the Decision Tree classifier showcased exceptional performance in accurately classifying players into specific rating clusters. With high accuracy, precision, recall, and F1 scores all exceeding 0.99, the classifier demonstrated robust predictive capabilities, enabling precise assignment of players to performance categories. This highlights the utility of decision tree algorithms in player assessment and strategic decision-making for coaches and analysts.

Conclusion

In conclusion, this project underscores the efficacy of machine learning techniques in accurately predicting the overall rating scores of football players. Through regression modeling, clustering algorithms, and decision tree classifiers, valuable insights have been gleaned into player performance profiles, offering coaches and analysts actionable intelligence for talent evaluation and strategic decision-making. Looking ahead, future work could involve feature engineering to incorporate additional player attributes, model optimization for improved performance, and the exploration of ensemble methods for enhanced accuracy. Additionally, the development of systems for real-time prediction and careful consideration of ethical implications are crucial for advancing the field of sports analytics responsibly. By addressing these areas and leveraging advancements in machine learning research, we can continue to refine predictive models and contribute to the ongoing evolution of football player performance prediction in a meaningful manner.

References

ABİDİN, DİDEM. 2021. "A Case Study on Player Selection and Team Formation in Football with Machinelearning." *Turkish Journal of Electrical Engineering and Computer Sciences* 29 (3): 1672–91. https://doi.org/10.3906/elk-2005-27.

Alanazi, Abdullah. 2022. "Using Machine Learning for Healthcare Challenges and Opportunities." *Informatics in Medicine Unlocked* 30 (January): 100924. https://doi.org/10.1016/j.imu.2022.100924.

Apostolou, Konstantinos, and Christos Tjortjis. 2019. "Sports Analytics Algorithms for Performance Prediction." In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–4. https://doi.org/10.1109/IISA.2019.8900754.

Boden, Margaret A. 2018. *Artificial Intelligence: A Very Short Introduction*. Oxford University Press.

Callahan, Alison, and Nigam H. Shah. 2017. "Chapter 19 - Machine Learning in Healthcare." In *Key Advances in Clinical Informatics*, edited by Aziz Sheikh, Kathrin M. Cresswell, Adam Wright, and David W. Bates, 279–91. Academic Press. https://doi.org/10.1016/B978-0-12-809523-2.00019-4.

Chandola, Varun, Arindam Banerjee, and Vipin Kumar. 2009. "Anomaly Detection: A Survey." *ACM Computing Surveys* 41 (3): 15:1-15:58. https://doi.org/10.1145/1541880.1541882.

Charbuty, Bahzad, and Adnan Abdulazeez. 2021. "Classification Based on Decision Tree Algorithm for Machine Learning." *Journal of Applied Science and Technology Trends* 2 (01): 20–28. https://doi.org/10.38094/jastt20165.

Cooil, Bruce, Lerzan Aksoy, and Timothy L. Keiningham. 2008. "Approaches to Customer Segmentation." *Journal of Relationship Marketing* 6 (3–4): 9–39. https://doi.org/10.1300/J366v06n03_02.

Duarte, Vannessa, Sergio Zuniga-Jara, and Sergio Contreras. 2022. "Machine Learning and Marketing: A Systematic Literature Review." *IEEE Access* 10: 93273–88. https://doi.org/10.1109/ACCESS.2022.3202896.

Hadley, Lawrence, Marc Poitras, John Ruggiero, and Scott Knowles. 2000. "Performance Evaluation of National Football League Teams." *Managerial and Decision Economics* 21 (2): 63–70. https://doi.org/10.1002/1099-1468(200003)21:2<63::AID-MDE964>3.0.CO;2-O.

Kok, Joost N., Egbert J. Boers, Walter A. Kosters, Peter Van der Putten, and Mannes Poel. 2009. "Artificial Intelligence: Definition, Trends, Techniques, and Cases." *Artificial Intelligence* 1 (270–299): 51.

Manish., S, Vandana Bhagat, and RM Pramila. 2021. "Prediction of Football Players Performance Using Machine Learning and Deep Learning Algorithms." In *2021 2nd International Conference for Emerging Technology (INCET)*, 1–5. https://doi.org/10.1109/INCET51464.2021.9456424.

Masood Ahmed. n.d. "Football Players Data." [object Object]. Accessed May 12, 2024. https://doi.org/10.34740/KAGGLE/DSV/6960429.

McIntosh, Sam, Stephanie Kovalchik, and Sam Robertson. 2019. "Comparing Subjective and Objective Evaluations of Player Performance in Australian Rules Football." *PLOS ONE* 14 (8): e0220901. https://doi.org/10.1371/journal.pone.0220901.

Ngai, Eric W. T., and Yuanyuan Wu. 2022. "Machine Learning in Marketing: A Literature

Review, Conceptual Framework, and Research Agenda." *Journal of Business Research* 145 (June): 35–48. https://doi.org/10.1016/j.jbusres.2022.02.049.
Pantzalis, Victor Chazan, and Christos Tjortjis. 2020. "Sports Analytics for Football League Table and Player Performance Prediction." In *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA*, 1–8. https://doi.org/10.1109/IISA50023.2020.9284352.
Pyda, Srinivas, and Srinivas Kareenhalli. 2018. "Mathematics and Machine Learning." In *Mathematics and Computing*, edited by Debdas Ghosh, Debasis Giri, Ram N. Mohapatra, Kouichi Sakurai, Ekrem Savas, and Tanmoy Som, 135–53. Singapore: Springer. https://doi.org/10.1007/978-981-13-2095-8_12.
Rommers, Nikki, Roland Rössler, Evert Verhagen, Florian Vandecasteele, Steven Verstockt, Roel Vaeyens, Matthieu Lenoir, Eva D'Hondt, and Erik Witvrouw. 2020. "A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players." *MEDICINE AND SCIENCE IN SPORTS AND EXERCISE* 52 (8): 1745–51. https://doi.org/10.1249/mss.0000000000002305.
Sheikh, Haroon, Corien Prins, and Erik Schrijvers. 2023. "Artificial Intelligence: Definition and Background." In *Mission AI: The New System Technology*, edited by Haroon Sheikh, Corien Prins, and Erik Schrijvers, 15–41. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-21448-6_2.
Sorzano, C. O. S., J. Vargas, and A. Pascual Montano. 2014. "A Survey of Dimensionality Reduction Techniques." arXiv. https://doi.org/10.48550/arXiv.1403.2877.
Suthaharan, Shan. 2016. "Support Vector Machine." In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, edited by Shan Suthaharan, 207–35. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-7641-3_9.
Yiğit, Ahmet Talha, Barış Samak, and Tolga Kaya. 2020. "Football Player Value Assessment Using Machine Learning Techniques." In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making*, edited by Cengiz Kahraman, Selcuk Cebi, Sezi Cevik Onar, Basar Oztaysi, A. Cagri Tolga, and Irem Ucal Sari, 289–97. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-23756-1_36.
Yousri, Ramez, and Gehad Hany. 2022. "Artificial Intelligence: An Overview." In *The International Undergraduate Research Conference*, 6:1–7. The Military Technical College.